

# Scaling Up of Action Repertoire in Linguistic Cognitive Agents

Vadim Tikhanoff and Angelo Cangelosi

Adaptive Behaviour & Cognition, University of Plymouth, Plymouth PL4 8AA, UK, vadim.tikhanoff@plymouth.ac.uk;  
acangelosi@plymouth.ac.uk

José F. Fontanari

Universidade de São Paulo, São Carlos, Brazil, fontanari@ifsc.usp.br

Leonid I. Perlovsky

Harvard University, Cambridge MA and The Air Force Research Laboratory, SN, Hanscom, MA

Leonid.Perlovsky@hanscom.af.mil

**Abstract** — We suggest the utilization of the Modeling Field Theory (MFT) to deal with the combinatorial complexity problem of language modeling in cognitive robotics. In new simulations we extend our previous MFT model of language to deal with the scaling up of the robotic agent's action repertoire. Simulations are divided into two stages. First agents learn to classify 112 different actions inspired by an alphabet system (the semaphore flag signaling system). In the second stage, agents also learn a lexical item to name each action. At this stage the agents will start to describe the action as a "word" comprised of three letters (consonant - vowel - consonant). The results of the simulations demonstrate that: (i) agents are able to acquire a complex set of actions by building sensorimotor concept-models; (ii) agents are able to learn a lexicon to describe these objects/actions through a process of cultural learning; and (iii) agents learn actions as basic gestures in order to generate composite actions.

## 1. INTRODUCTION

Recent research in autonomous cognitive systems has focused on the close integration (grounding) of language with perception and other cognitive capabilities [1]-[4]. This approach is based on the important process of "grounding" the agent's lexicon directly into its own internal representations. Agents learn to name entities, individual and states whilst they interact with the world and build sensorimotor representations of it. For example Steels [5] studied the emergence of shared languages in group of autonomous cognitive robotics that learn categories of object shapes and colors. Cangelosi and collaborators analyzed the emergence of syntactic categories in lexicons supporting navigation [6] and object manipulation tasks [7, 8] in populations of simulated agents and robots.

Current grounded agent and robotic approaches have their own limitations, in particular for the scaling up of the agents' lexicon since they can only use few tens of lexical entries (see [5]) and can deal with a limited set of syntactic categories (e.g. nouns and verbs in [6]). This is mostly due to the use of computational intelligent techniques (e.g. neural networks, rule systems) subject to combinatorial complexity (CC). The issue of scaling up and CC in cognitive systems has been recently addressed by [9]. In linguistic systems, CC refers to the hierarchical combinations of bottom-up perceptual and linguistic signals and top-down internal concept-models of objects, scenes and other complex meanings. Perlovsky proposed the Modeling Field Theory (MFT) as a new method for overcoming the exponential growth of CC in computational intelligent techniques currently used in cognitive systems design. MFT uses fuzzy dynamic logic to avoid CC and computes similarity measures between internal concept-models and the perceptual and linguistic signals. More recently, Perlovsky [10] has suggested the use of MFT specifically to model linguistic abilities. By using concept-models with multiple sensorimotor modalities, a MFT system can integrate language-specific signals with other internal cognitive representations.

Perlovsky's proposal to apply MFT in the language domain is highly consistent with the grounded approach to language modeling discussed above. That is, both accounts are based on the strict integration of language and cognition. This permits the design of cognitive systems that are truly able to "understand" the meaning of words being used by autonomously linking the linguistic signals to the internal concept-models of the word constructed during the sensorimotor interaction with the environment. The combination of MFT systems with grounded agent simulations will permit the overcoming of the CC problems currently faced in grounded agent models and scale up the lexicons in terms of high number of lexical entries and syntactic categories.

In this paper we propose the utilization of the Modeling Field Theory (MFT) to deal with the combinatorial complexity problem of language modeling. MFT aims at overcoming such limitations by dynamic logic learning of

lower-level signals (e.g., inputs, bottom-up signals) with hierarchies of higher-level concept-models (e.g. internal representations, categories/concepts, top-down signals). This is the case of language, which is characterized by the hierarchical organization of underlying cognitive models. Modeling Field Theory may be viewed as an unsupervised learning algorithm whereby a series of concept-models adapt to the features of the input stimuli via gradual adjustment dependent on the fuzzy similarity measures.

In this paper we present an integration of the Modeling Field Theory algorithm for the classification of objects with a model of the acquisition of language in cognitive robotics. We will further extend our previous modified version of the MFT algorithm [11] to deal with the scaling up of the robotic agent's action repertoire. The new extended MFT model will be presented in Section 2. Simulation setups and results are reported in Section 3.

## 2. MATHEMATICAL FRAMEWORK

We consider the problem of categorizing  $N$  objects  $i=1, \dots, N$ , each of which characterized by  $d$  features  $e=1, \dots, d$ . These features are represented by real numbers  $O_{ie} \in (0,1)$  - the input signals. Accordingly, we assume that there are  $M$   $d$ -dimensional concept-models  $k=1, \dots, M$  described by real-valued fields  $S_{ke}$ , with  $e=1, \dots, d$  as before, that should match the object features  $O_{ie}$ . Since each feature represents a different property of the object as, for instance, color, smell, texture, height, etc. and each concept-model component is associated to a sensor sensitive to only one of those properties, we must, of course, seek for matches between the same component of objects and concept-models. Hence it is natural to define the following partial similarity measure between object  $i$  and concept-model  $k$  [9]

$$l(i|k) = \prod_{e=1}^d (2\pi\sigma_{ke}^2)^{-1/2} \exp[-(O_{ie} - S_{ke})^2 / 2\sigma_{ke}^2] \quad (1)$$

where, at this stage, the fuzziness  $\sigma_{ke}$  are parameters given a priori. The goal is to find an assignment between models and objects such that the global (log) similarity

$$L = \sum_i \log \sum_k l(i|k) \quad (2)$$

is maximized. This maximization can be achieved using the MFT mechanism of concept formation which is based on the following dynamics for the modeling field components

$$dS_{ke}/dt = \sum_i f(k|i) [\partial \log l(i|k) / \partial S_{ke}] \quad (3)$$

where

$$f(k|i) = l(i|k) / \sum_{k'} l(i|k') \quad (4)$$

are the fuzzy association variables which give a measure of the correspondence between object  $i$  and concept  $k$  relative to all other concepts  $k'$ . This quantity can be viewed as (adaptive) neural weights that yield the strength of the association between input and concepts. Using the explicit expression for the similarity measure, Eq. (1), the dynamic equations become

$$dS_{ke}/dt = \sum_i f(k|i) (O_{ie} - S_{ke}) / \sigma_{ke}^2 \quad (5)$$

for  $k=1, \dots, M$  and  $e=1, \dots, d$ . From Eq. (5) it becomes clear that the fuzzy association variables are responsible for the coupling of the equations for the different modeling fields and, even more importantly for our purposes, for the coupling of the distinct components of a same field. In this sense, the categorization of multi-dimensional objects is not a straightforward extension of the one-dimensional case because new dimensions should be associated with the appropriate models [11]. This nontrivial interplay between the field components will become clearer in the discussion of the simulation results.

It can be shown that the dynamics (5) always converges to a (possibly local) maximum of the similarity  $L$  [9], but by properly adjusting the fuzziness  $\sigma_{ke}$  the global maximum often can be attained. A salient feature of dynamic logic is a match between parameter uncertainty and fuzziness of similarity. In what follows we decrease the fuzziness during the time evolution of the modeling fields according to the following prescription

$$\sigma_{ke}^2(t) = \sigma_a^2 \exp(-\alpha t) + \sigma_b^2 \quad (6)$$

with  $\alpha = 5 \times 10^{-4}$ ,  $\sigma_a = 1$  and  $\sigma_b = 0.03$ . Unless stated otherwise, these are the parameters we will use in the forthcoming analysis.

## 3. SIMULATIONS

In this section we will report results from three computational experiments. Initially they will be aimed at a simple scaling up of the agent's action repertoire using multi-dimension features. In the second simulation we will demonstrate the correct classification of the input object though the dynamic introduction of the lexicon feature. The third simulation will concentrate on breaking down the actions into basic gestures in order to generate composite actions. To facilitate the presentation of the results, we will interpret both the object feature values and the modeling fields as  $d$ -dimensional vectors and follow the time evolution of the corresponding vector length

$$S_k = \sqrt{\sum_{e=1}^d (S_{ke})^2} / d \quad (7)$$

which should then match the object length

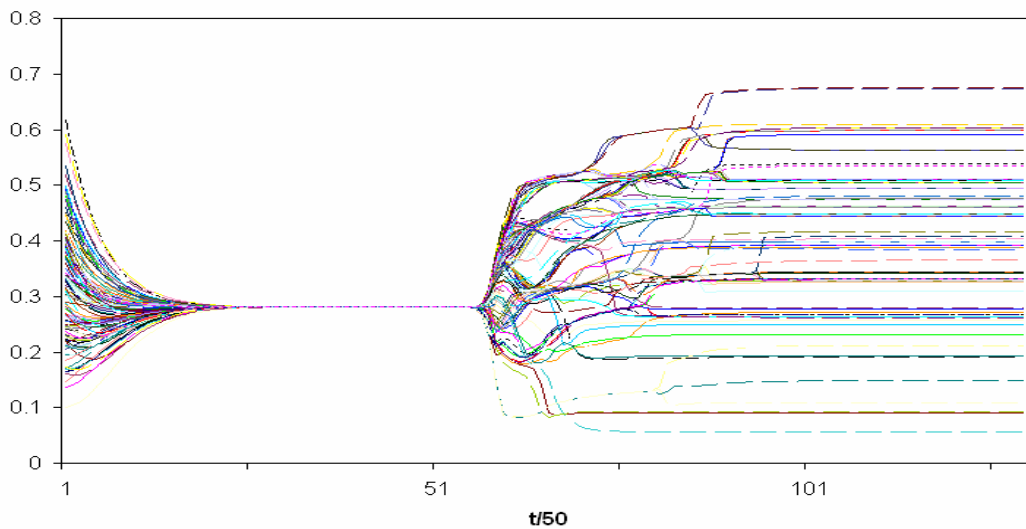
$$O_i = \sqrt{\sum_{e=1}^d (O_{ie})^2} / d.$$

**Simulation I: Classification and categorization of actions / building sensorimotor concept-models**

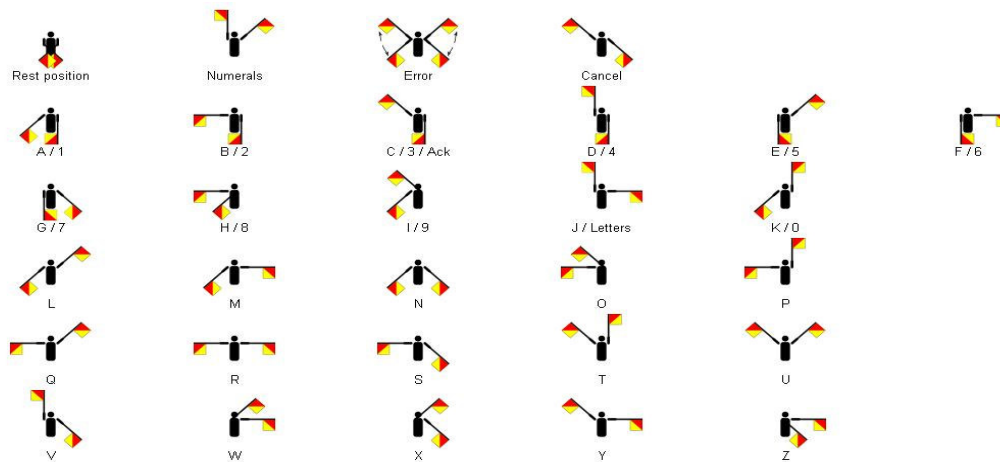
Let's first consider having 112 different actions, some inspired by an alphabet system (the semaphore flag signaling system, see Figure 2). We have collected data on the posture of robots using 6 features. The object input data consist of the 6 angles of each, left arm and right arm joints (shoulder, upper arm and elbow). The agents first have to learn to classify these actions; at this stage we are using a multi-dimensional MFT algorithm with 112 fields randomly initialized. Figure 1 shows that the model is

able to correctly identify the different actions. The time is presented in units of the time step  $h$  of Euler's algorithm used to solve the coupled set of dynamic equations. Although the simulation initially dealt with 112 actions the MFT algorithm was able to categorize to approximately 95% successful matching. Therefore there was a slight reduction in the number of completed actions. Figure 3 shows our system consisting of two simulated agents - teacher and learner - embedded within a virtual simulated environment (using Open Dynamic Engine). In respect to equation (1), in this experiment  $M=N=112$  and  $d = 6$  features.

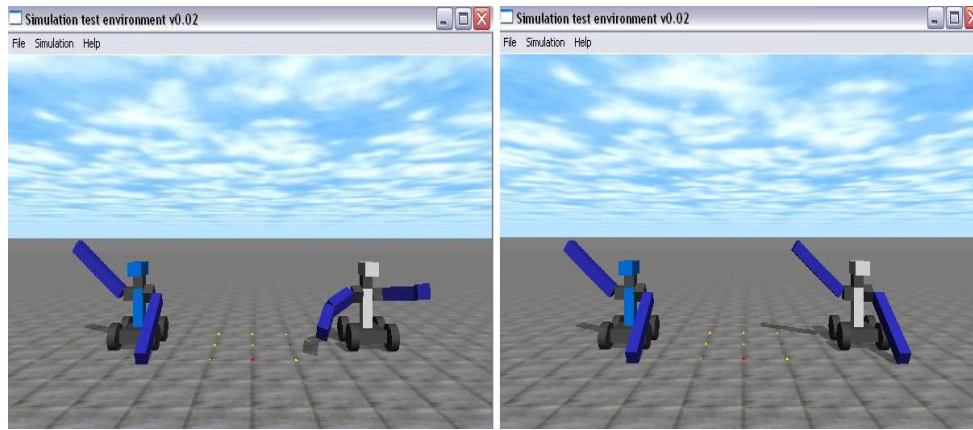
**Classification and categorization of actions**



**Figure 1** - Time evolution of the fields with 6 features being used as input: 112 different actions



**Figure 2** - Few examples of type of behavior used for the classification and categorization of actions. (Here the semaphore alphabet)



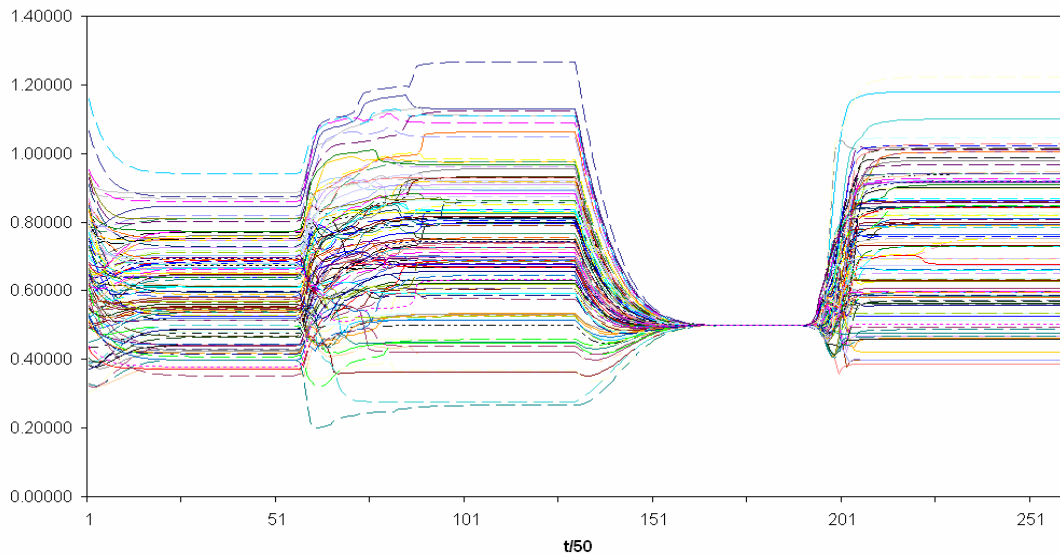
**Figure 3** - Teacher and learner before (left) and after (right) the action is learnt.

### Simulation II: Incremental Feature – lexicon acquisition

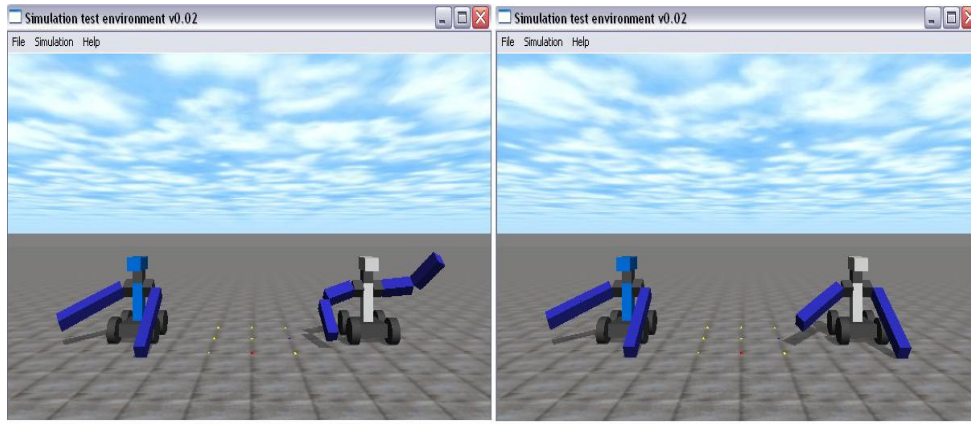
In the first simulation we have proposed the use of the multi-dimensional MFT in order to categorize 112 different actions. At this stage we wanted to explore the integration of language and cognition in cognitive robotic studies. Here we extend the multi-dimensional MFT algorithm, used in Simulation 1, to enable the agents to learn a lexical item to name each previous action. After performing the action, the agents will start to describe it as three letters words (consonant – vowel – consonant; for example: “XUM”, “HAW”, “RIV”, etc.). Each letter uses

two features therefore each word is represented by 6 additional features. Each word is unique to the action performed. This phonetic feature is dynamically added immediately after the action. At timestep 12500, (half of the training time) both features are considered when computing the fuzzy similarities. From timestep 12500, the dynamics of the  $\sigma_2$  fuzziness value is initialized, following equation (6), whilst  $\sigma_1$  continues its decrease pattern started at timestep 0. Results in Fig. 4 show that the model is able to categorize an action and assign a ‘word’ to this action. In this experiment  $d = 12$  comprising of the robot and phonetic features.

### : Incremental Feature – lexicon acquisition



**Figure 4** - Time evolution of the fields using as input the action and phonetic feature: 112 different actions + 112 words



“XUM”

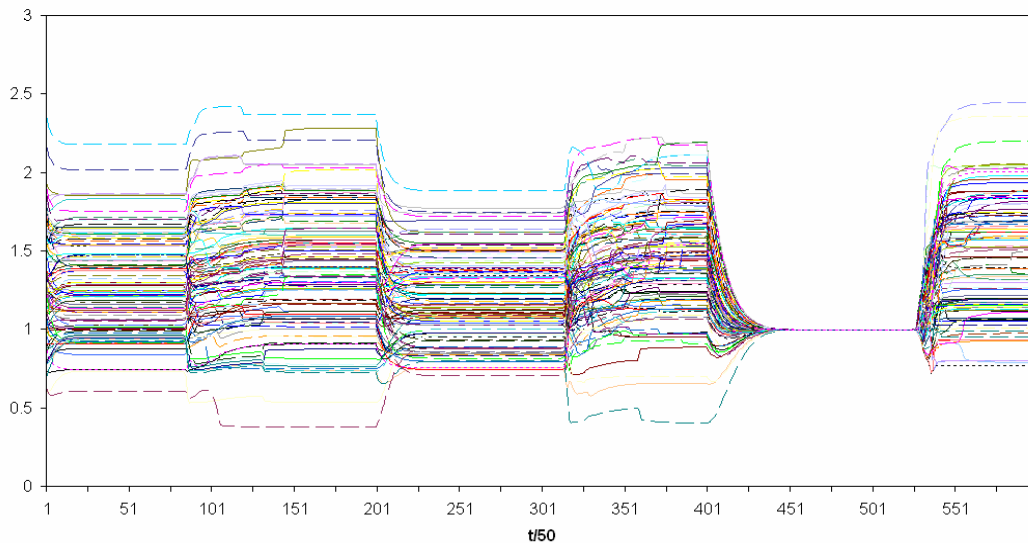
**Figure 5-** Teacher and learner before action is learnt and after with the addition of the ‘word’. For visualization purposes, the word is added on the image.

**Simulation III: Progressive learning of basic gestures into composite actions**

The previous simulations consisted of learning actions or a combination of actions and words. In this final simulation we take a step backwards in the categorization of actions and break down the action into basic gestures. Before learning a complete action we are interested in the systematic breakdown of actions into individual gestures, that is to say for example a two-handed action would be broken down into two single handed-actions and analyzed as individual steps in the process of a compound action. As an extension to the previous simulations, each feature is added dynamically. The simulation starts with the left-handed action. Then at timestep 10000 (1/3rd of the

simulation run) we consider the right-handed action, using the same dynamics of the fuzziness values as for simulation II, and finally at timestep 20000 we consider the phonetic feature. Figure 6 shows that the model is able to dynamically adapt to compound action associated with the word generation.

**Progressive learning of basic gestures into composite actions**



**Figure 6:** Time evolution of the fields using as input the composite action and phonetic feature: 112 different composite actions + 112 words

## 4. CONCLUSION

In this paper we presented an integration of the Modeling Field Theory algorithm for the classification of objects with a model of the acquisition of language in cognitive robotics. In new simulations we have applied and extended our previous modified version of the MFT algorithm to deal with the scaling up of the robotic agent's action repertoire. The various simulations showed that (i) agents are able to acquire a complex set of actions by building sensorimotor concept-models; (ii) agents are able to learn a lexicon to describe these objects/actions through a process of cultural learning. (iii) agents learn actions as basic gestures in order to generate composite actions.

Future work will look at the further development of the MFT algorithm to allow a more implicit link between action representations and lexicons and the learning of meaning-word pairs. In addition, we are going to test the above simulation model in a hardware robotic platform.

## ACKNOWLEDGMENTS

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grants number FA8655-05-1-3060 and FA8655-05-1-3031. The U.S. Government is authorized to reproduce and distribute reprints for Government purpose notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

## REFERENCES

- [1] A. Cangelosi and T. Riga, "An Embodied Model for Sensorimotor Grounding and Grounding Transfer: Experiments with Epigenetic Robots", *Cognitive Science*, **30**, 1-17 (2006).
- [2] D. Pecher and R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking*, Cambridge: Cambridge University Press, 2005.
- [3] S. Harnad, "The symbol grounding problem", *Physica D* **42**, 335-346 (1990).
- [4] A. Cangelosi, G. Bugmann and R. Borisyuk (Eds.), *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*, Singapore: World Scientific, 2005.

- [5] L. Steels, "Evolving grounded communication for robots", *Trends in Cognitive Sciences*, **7**, 308-312 (2003).
- [6] A. Cangelosi, "Evolution of communication and language using signals, symbols and words", *IEEE Transactions on Evolutionary Computation* **5**, 93-101 (2001).
- [7] D. Marocco, A. Cangelosi and S. Nolfi, "The emergence of communication in evolutionary robots", *Philosophical Transactions of the Royal Society of London – A* **361**, 2397-2421 (2003).
- [8] A. Cangelosi, E. Hourdakis and V. Tikhonoff, "Language acquisition and symbol grounding transfer with neural networks and cognitive robots" *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.
- [9] L. I. Perlovsky, *Neural Networks and Intellect: Using Model-Based Concepts*, Oxford: Oxford University Press, 2001.
- [10] L. I. Perlovsky, "Integrating language and cognition," *IEEE Connections* **2**, 8-13 (2004).
- [11] V. Tikhonoff, J.F. Fontanari, A. Cangelosi and L.I. Perlovsky, "Language and Cognition Integration Through Modeling Field Theory: Category Formation for Symbol Grounding", *Lecture Notes in Computer Science* **4131**, 376-385, 2006.